Exhibit A

The Computer-Aided Analysis of Crossover Studies
Gerard E. Dallal, Ph.D.

Abstract

This note describes the computer-aided analysis of two treatment, two-period crossover studies. All participants are given both treatments. Half of the subjects receive the treatments in one order, the others receive the treatments in the reverse order. SAS and SYSTAT command language is given for the analysis of such trials.

Introduction

Most studies of two treatments--A and B, say--are parallel groups studies, so-called because the treatments are studied in parallel. One group of subjects receives only treatment A, the other group receives only treatment B. At the end of the study, the two groups are compared on some quantitative outcome measure (a final value of some marker, a change from baseline, or the like), most often by using a t test for independent samples.

It takes little experience with parallel group studies to recognize the potential for great gains in efficiency if each subject could receive both treatments. The comparison of treatments would no longer be contaminated by the variability between subjects since the comparision is carried out within each individual.

If all subjects received the two treatments in the same order, observed differences between treatments would be confounded with any other changes that occur over time. In a study of the effect of treatments on cholesterol levels, for example, subjects might change their diet and exercise behavior for the better as a result of heightened awareness of health issues. This would likely manifest itself as a decrease in cholesterol levels over the later portion of the study and might end up being attributed to the second treatment.

The two treatment, two-period crossover study seeks to overcome this difficulty by having half of the subjects receive treatment A followed by treatment B while the other half receive B followed by A. The order of administration is incorporated into the formal analysis. In essence, any temporal change that might favor B over A in one group will favor A over B in the other group and cancel out of the treatment comparison.

Even though crossover studies are conceptually quite simple, the literature is difficult to read for many reasons.

1. Terminology and notation varies from author to author, making it difficult to compare discussions.

Reference	Terminology		
[2]	TREATMENT	PERIOD	TREATMENT*PERIOD
[3]	TREATMENT	PERIOD	SEQUENCE
[6]	TREATMENT	SEQUENCE*TREATMENT	SEQUENCE

Complicated notational devices are introduced to describe simple comparisons among four cell means.

- In Grizzle [1], y_{ijk} represents the response of subject j to treatment k applied at time period i.
- In Hills and Armitage [2], y_i represents the response of a subject in period i. The treatment is implied by context. The difference between treatments X and Y for group A is $d_A = y_1 - y_2$, while the difference between treatments X and Y for group B is $d_B = y_2 - y_1$.
- In Fleiss [3], X_j represent the response in period j of a subject who receives the treatments in the first order and Y_j represent the response in period j of a subject who receives the treatments in the second order.

2. Factors such 'time period' and 'sequence in which the treatments are given' are easily confused when reduced to the one word labels required by printed tables or computer programs.

3. The mathematical theory for cross-over studies was developed before the ready availabilty of computers, so practical discussions concentrated on methods of analysis that could be carried out by hand. Because the basic crossover involves only two treatments and two periods, most authors give the analysis in terms of t tests. Virtually all general purpose computer programs analyze crossover studies as special cases of repeated measure analysis of variance and give the results in terms of F tests. The two approaches are algebraically equivalent, but the difference in appearance makes it difficult to reconcile computer output with textbooks and published papers.

4. Many published discussions and examples are incorrect. Grizzle [1] gave an incorrect analysis of studies with unequal numbers of subjects in each sequence group. Nine years later, Grizzle [4] corrected the formula for sums of squares for treatments. After an additional eight years elapsed, Grieve [5] noted, "Although . . . it should be clear that the period sum of squares . . . is also incorrect, the analysis put forward in Grizzle [1,4] still appears to be misleading people. I know of three examples of computer programs, written following Grizzle's analysis, with incorrect period and error sums of squares."

   Grizzle's flawed analysis continues to muddy the waters. In their otherwise excellent book, "intended for everyone who analyzes data," Milliken and Johnson [6] present the flawed analysis in Grizzle [1]; Grizzle [4] is not referenced.

The Crossover Study

In this discussion, the design factors will be denoted

- treatment (TREATMENT), with levels 'A' and 'B',
- time period (PERIOD), with levels '1' and '2',
- sequence group (GROUP), with levels 'A then B' and 'B then A'.

Some authors prefer SEQUENCE to GROUP. There are two powerful reasons for using GROUP. First, the word GROUP is unambiguous. It implies differences between subjects, and there is only one way subjects differ--in the order in which they receive the two treatments. Confusion over SEQUENCE/PERIOD/ORDER is eliminated. Second, the error term for testing the significance of the GROUP factor is different from the error term for testing PERIOD and TREATMENT, as is true of any repeated measures study with between- and within-subjects factors. The label GROUP helps keep this in mind.

The four observed cell means

Group 1:	A in period 1	B in period 2
Group 2:	B in period 1	A in period 2

will be denoted

x_1	x_2
x_3	x_4

While this prescription introduces yet another set of notation, here the notation is neutral--no attempt has been made to describe the experiment through the notation. When discussing a set of four numbers, this neutrality proves a virtue rather than a vice.

Statistical Details

TREATMENTS are compared by combining the difference between A and B from within each group, specifically

$$((x_1 - x_2) + (x_4 - x_3)) / 2 .$$

PERIODS are compared by looking at the difference between the measurements in period 1 and those made in period 2

$$((x_1 + x_3) - (x_2 + x_4)) / 2 .$$

If period effects are present, they do not influence the comparison of treatments. A period 1 effect appears in the treatment comparisons as part of x_1 and x_3 and cancels out of the treatment difference, while a period 2 effect appears as part of x_2 and x_4 and cancels out, as well. Similarly, treatment effects do not influence the comparison of time periods. Treatment A appears in x_1 and x_4 and cancels out of the PERIODS effect, while treatment B appears as part of x_2 and x_3 and cancels out, too.

Aliasing

If crossover studies were full-factorial designs (with factors GROUP, TREATMENT, and PERIOD), it would be possible to evaluate not only the main effects, but also the GROUP*TREATMENT, PERIOD*TREATMENT, GROUP*PERIOD, and PERIOD*TREATMENT*GROUP interactions. However, crossover studies are not full factorial designs. Not all combinations of factors appear in the study (there is no GROUP='A then B', PERIOD='1', TREATMENT='B' combination, for example). Because only four combinations of the three factors are actually observed, main effects are confounded with two-factor interactions, that is, **each estimate of a main effect also estimates a two-factor interaction**.

As an illustration, notice that the difference between GROUPS is estimated by comparing the two means for group 1 to the two means for group 2, that is,

Group 1   Group 2

-

$$(x_1 + x_2) \qquad\qquad (x_3 + x_4)$$

Now consider the PERIOD*TREATMENT interaction, which measures how the difference between treatments change over time. The interaction is estimated by

	Period 1		Period 2
	A - B		A - B
	$(x_1 - x_3)$	-	$(x_4 - x_2)$

But this is the estimate of the GROUP effect. Thus, GROUP and PERIOD*TREATMENT are confounded. They are *aliases*, two names for the same thing. In the two treatment, two period crossover study, each main effect is confounded with the two-factor interaction involving the other factors.

Effect	Alias
TREATMENT	GROUP*PERIOD
PERIOD	GROUP*TREATMENT
TREATMENT*PERIOD	GROUP

If one of the main effects is significant, it is impossible to tell whether the effect, its alias, or both are generating the significant result. One could argue that there is no reason to expect a significant effect involving GROUP because subjects are assigned to GROUPS at random. Therefore, a significant GROUP effect should be interpreted as resulting from a PERIOD*TREATMENT interaction and not from a difference between GROUPS. For similar reasons, a significant PERIOD effect is not considered to be the result of a GROUP*TREATMENT interaction, nor is a significant TREATMENT effect to be the result of a GROUP*PERIOD interaction.

Carryover Effects

Carryover (or residual) effects occur when the effect of a treatment given in the first time period persists into the second period and distorts the effect of the second treatment. Carryover effects will cause the difference between the two treatments to be different in the two time periods, resulting in a significant TREATMENT*PERIOD interaction. Thus TREATMENT*PERIOD is not only an alias for GROUP, it is also another way of labelling CARRYOVER effects.

When the TREATMENT*PERIOD interaction is significant, indicating the presence of carryover, a usual practice is to set aside the results of the second time period and analyze the first period only.

The Computer-Aided Analysis of Crossover Studies

Crossover designs are easily analyzed by any statistical program package, such as SAS (SAS Institute, Cary, NC) and SYSTAT (SPSS Inc., Chicago, IL), that can perform repeated measures analysis of variance.

Within each record, a subject's data can be ordered by either treatment or time period. Both arrangements for the data set given in Grizzle [1] are appended to the end of this note along with the appropriate SAS PROC GLM control language. For data ordered by TREATMENT, the test of treatments will be labelled TREATMENT, the test of treatment by period interaction (which is also the carryover effect) will be labelled GROUP, and the test of time periods will be labelled

GROUP*TREATMENT, in keeping with the list of aliases developed earlier. For data ordered by PERIOD the test of treatments will be labelled GROUP*PERIOD, the test of treatment by period interaction will be labelled GROUP, and the test of time periods will be labelled PERIOD.

It might seem more natural to arrange the data by TREATMENT. This has the advantage of having the treatment comparison labelled TREATMENT. If the data are arranged by PERIOD, however, it is easier to analyze only the data from the first period data if a significant PERIOD*TREATMENT interaction is found.

SYSTAT command language is similar. The instructions for data ordered by TREATMENT are

```
CATEGORY GROUP
MODEL A B = CONSTANT + GROUP/REPEATED, NAME = "Treat"
ESTIMATE
```

The instructions for data ordered by PERIOD are

```
CATEGORY GROUP
MODEL P1 P2 = CONSTANT + GROUP/REPEATED, NAME = "Period"
ESTIMATE
```

Practical Issues

"The intuitive appeal of having each subject serve as his or her own control has made the crossover study one of the most popular experimental strategies since the infancy of formal experimental design. Frequent misapplications of the design in clinical experiments, and frequent misanalyses of the data, motivated the Biometric and Epidemiological Methodology Advisory Committee to the U.S. Food and Drug Administration to recommend in June of 1977 that, in effect, the crossover design be avoided in comparative clinical studies except in the rarest instances." Fleiss [3, p. 263]

Despite the appeal of having each subject serve as his own control, crossover studies have substantial weaknesses, as well, even beyond the possibility of carryover effects mentioned earlier. Because subjects receive both treatments, crossover studies requires subjects to be available for twice as long as would be necessary for a parallel groups study and perhaps even longer, if a washout period is required between treatments. Acute problems might be gone before the second treatment is applied. A washout period between the two treatments might minimize the effects of the carryover, but this will not be feasible for treatments like fat soluble vitamin supplements that can persist in the body for months.

On the other hand, some features of the crossover may make the design preferable to a parallel groups study. In certain cases, volunteers might be willing to participate only if they receive a particular treatment. The crossover insures that each subject will receive both treatments.

References

1. Grizzle JE. The two-period change-over design and its use in clinical trials. Biometrics 21, 467-480 (1965).
2. Hills M, Armitage P. The two-period cross-over clinical trial. British Journal of Clinical Pharmacology 8, 7-20 (1979).
3. Fleiss JL. The Design and Analysis of Clinical Experiments. John Wiley & Sons, Inc., New York (1986).
4. Grizzle JE. Correction. Biometrics 30, 727 (1974).
5. Grieve AP. Correspondence: The two-period changeover design in clinical trials. Biometrics 38, 517 (1982).
6. Milliken GA, Johnson DE. Analysis of Messy Data. Van Nostrand Reinhold Co., New York (1984).

```
                Data from Grizzle [1] arranged for
                 analysis by using SAS's PROC GLM

      <data arranged                       <data arranged
         by treatment>                        by period>

       DATA;                               DATA;
         INPUT GROUP A B;                    INPUT GROUP P1 P2;
         CARDS;                              CARDS;
       1  0.2  1.0                         1  0.2  1.0
       1  0.0 -0.7                         1  0.0 -0.7
       1 -0.8  0.2                         1 -0.8  0.2
       1  0.6  1.1                         1  0.6  1.1
       1  0.3  0.4                         1  0.3  0.4
       1  1.5  1.2                         1  1.5  1.2
       2  0.9  1.3                         2  1.3  0.9
       2  1.0 -2.3                         2 -2.3  1.0
       2  0.6  0.0                         2  0.0  0.6
       2 -0.3 -0.8                         2 -0.8 -0.3
       2 -1.0 -0.4                         2 -0.4 -1.0
       2  1.7 -2.9                         2 -2.9  1.7
       2 -0.3 -1.9                         2 -1.9 -0.3
       2  0.9 -2.9                         2 -2.9  0.9
       ;                                   ;

       PROC GLM;                           PROC GLM;
         CLASS GROUP;                        CLASS GROUP;
         MODEL A B = GROUP/NOUNI;            MODEL P1 P2 = GROUP;
         REPEATED TREAT 2/SHORT;             REPEATED PERIOD 2/SHORT;
```